A Survey on Mental State Analysis using Natural Language Processing

Satya Verma[1], Abhishek Shrivastava[2]
Assistant Professor[1], Assistant Professor[2]
Mats University[1], NIT Raipur[2]

Abstract: Human beings are full of emotions & the mental state of human beings is varied depending on the surrounding and situation. As the situation changes so as the mood of the person also changes. Being a human we are also able to recognize someone's emotion by their facial expressions and their wordings. But when try to recognize someone's emotion or sentiments in the written material, than it is a tedious job. At the same time if we want to simulate the same task by the machine then it is a difficult one. In this research proposal the emotion detection in text based documents are discussed. The phenomenon uses the concept of Natural Language Processing (NLP). In literature several kinds of emotion/sentiment identification has been proposed (like sentiment analysis in movies, news articles, social network etc). In this paper section 1 gives the brief introduction, section 2 describes the literature of the relevant work, section 3 gives the comparative analysis of the various methods, and section 4 gives the future scope with conclusion.

Index Terms - NLP, opinion mining, POS tagger, and corpus.

I. INTRODUCTION

Emotion is a very basic characteristic of every human being and we don't need much expertise to sense the other person's emotion. This emotion sensing can be done in numerous ways like by someone's facial expression, someone's speech or someone's act. There are various kinds of emotion, which a person can experience. Yu Zhang at el. Classified human emotions into 12 categories. They are: happy, sad, fearful, disgusted, angry, surprised, love, expectant, nervous, regretful, praiseful, shy [1]. Well it is not required to consider all kinds of emotions rather it totally depends on the application area in which we want to explore the emotion sensing.

Now a day's researchers are paying more attention towards processing of written text. While processing written text on machine, it all comes under NLP. Applications for processing large amounts of texts require NLP expertise. Some requirements are text classification into categories, indexing and searching of large texts, automatic translation, speech understanding, information extraction, automatic summarization, question answering, knowledge acquisition, and text generations [2].

Textual input for emotion/sentiment detection can be obtained from many sources such as text book, e-mails, messages, news articles, blogs, and social networks depending on the required text. Many researchers analyzed the emotion/sentiment of the textual documents in numerous fields. Review of news article, business article or movies is such kind of tasks in which we are having reports. And machine has to decide whether the textual reports are positive, negative, or neutral [3] [4]. Textual data can be subjective or objective. If one is trying to detect emotion/sentiment than it is better to use subjective text because it will provide the better way to detect information [9].

Sentiment classification is part of the opinion mining and refers to the identification of opinions and arguments in a text. There are two types of opinion –

1. Direct opinion and
2. Comparisons.

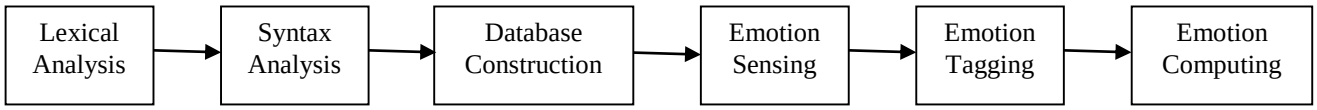
Direct opinions are the opinion expressing on the products, events, topics, and persons, etc. Comparisons express the similarities or differences between more than one object. The result of the opinions are used in business and organization by consultants, surveys and focused group, etc. and the individuals who are interested in other's opinions for purchasing a product or using a service, finding the opinions on political topics, in advertising placements and in opinion retrieval/search [5].

While processing natural languages, there are lots of challenges too. The mainly identified challenges are Data Collection, Choice of Features, Emotion Labeling, and Choice of Machine Learning Classifier [10]. Because we have to make the machine as much capable to recognize emotion what it exactly means. Sometimes it happens that a particular sentence has started with positive attitude and ended with a negative one or vice versa. So in this kind of situation the applied method should be able to detect correct intention of the author. Next section discusses about the methods used in the literature to perform the task of emotion analysis.

II. LITERATURE REVIEW

There are many researchers who are working in the field of NLP to detect sentiments of the document author. The intentions of all the researchers are to develop algorithms and detect exact emotion from the selected text. The program developed for NLP must have sound knowledge about the structure of the language including word knowledge and how they are combined into phrases and sentences. While processing spoken language the major steps of knowledge understanding are phonological analysis, morphological analysis, syntactic analysis, pragmatic analysis, and world analysis [2].

Yu Zhang at el. has proposed semi-automatic model for emotion detection for Chinese natural language processing. In the proposed model they followed the following steps [1]:



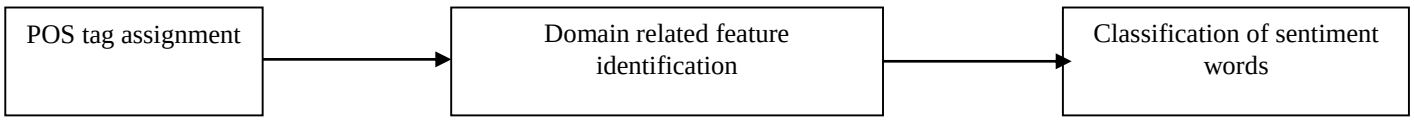
They used ICTCLAS lexical analyzer which is a Chinese lexical analyzer. Syntax analysis phase checks the correctness of the input sentence. Next a database had constructed with emotional information of Chinese words. They collected 4800 words from various dictionaries. Five levels have been defined (although they classified emotions into 12 categories):

Derogatory-> Negative-> Neutral-> Positive-> Commendatory

Next step defined is emotion sensing in which database is used to recognize the obscured category of the text. In lexical analysis sentences are splits into words and each word is tagged according to its category. According to above stated 5 levels, words are assigned tag from -2 to 2 depending upon their category. Emotion computing is done by considering all the words appeared in a sentence by simply calculating average of the tag values. The accuracy of the result was around 80% in the experimentation [1].

Sentiments are classified on the analysis of three levels – document level, sentence level, and feature level. In the document level and sentence level sentiment classifications assume that each document or sentence focus on a single object and contains only one opinion or opinion from a single opinion holder. Feature level sentiment classification, we need to identify and extract objects features that are commented on by the opinion holders, group features synonyms, and determine whether the opinions on the features are positive, negative or neutral on the software reviews.

Khin Phyu Shein at el. has proposed ontology based combination approach for sentiment classification [5]. They worked on feature level sentiment classification with three main parts:



POS tagger is used to assign POS tags to words in a sentence (such as tags for nouns, verbs, and adjective). Next domain ontology is used to get domain related features. The method which is used for this step is based on Formal Concept Analysis (FCA). The main characteristics of FCA are:

concepts are described by properties

the properties determine the hierarchy of the concepts

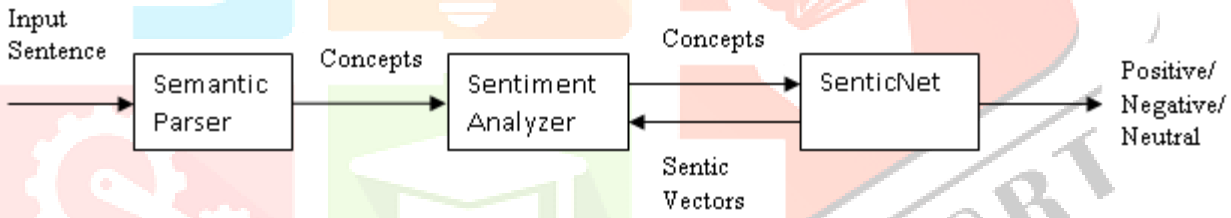
When the properties of different concepts are the same, then the concepts are the same.

FCA can be represented as triples (O, A, R)

Where O=finite set of object, A=finite set of attributes, and R=binary relation on O and A

Classification of sentiment words has done by SVM. Features are classified by linearly separated hyper plane with the binary classification. Then features are defined by labeling positive or negative [5].

Prashant Raina has proposed a method for review of news articles using sematic computing. In this paper, they presented an opinion-mining engine that exploits common-sense knowledge extracted from ConceptNet and Sentinel to perform sentiment analysis in news articles [3]. He used scenic computing for opinion mining. System proposed by the Prashant Raina is shown below:



The components of the model are –

A semantic parser which extracts concepts from a given sentence.

A local copy of the SenticNet database.

A sentiment analyzer, which consults the sentic vectors in SenticNet to classify each sentence as positive, negative or neutral.

Semantic parser extracts a set of candidate common-sense concepts from each sentence. Next sentiment analyzer matches the candidate concepts with sentic vectors in SenticNet. Thus a set of sentic vectors for each sentence describes the emotions present in it. Polarity measure is given by:

$$Polarity = \frac{P + 0.3 \times |T| - 0.3 \times |S| + 0.3 \times A}{1.9}$$

Here, P, T, S and A represent the pleasantness, attention, sensitivity and aptitude of the sentic vector respectively. Polarity score ranges from -1 to +1. And according to the polarity score, sentence is marked as positive, negative or neutral. The experimental result of this model provided 71.2% accuracy [3].

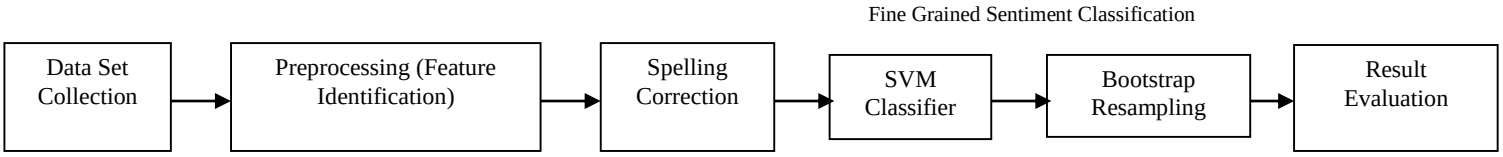
Yiming Zhao at el. has proposed optimization of Chinese text sentiment analysis for social network [6]. They did optimization of sentiment analysis method on micro-blog content (SAMC) in three steps – data structure optimization, query strategy optimization, and parallel optimization. In basic SAMC algorithm, the sentiment analysis of each piece of micro blog contains three main steps – sentence segmentation and word segmentation, polarity calculation of sub sentence, and polarity calculation of whole sentence.

$$\lambda(s_i) = \frac{1}{\min(1, N - i + 1)}, 1 \le i \le N$$

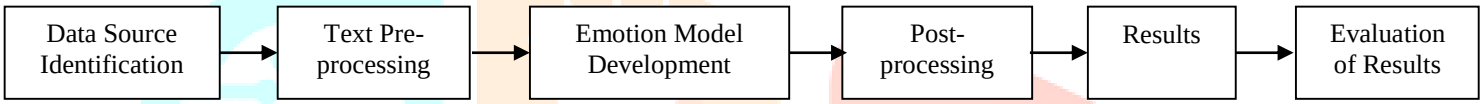
$$Polarity(S) = \sum_{i=1}^{N} [\lambda(s_i) \times p(s_i)]$$

Where S is a sentence, si is sub sentence, N is the number of sub-sentences in S, and i is the position of si in S. λ(si) and p(si) represent the position and polarity of sub-sentence si.

SAMC algorithm is optimized by:
Data structure optimization in which red black tree is used to store the vocabulary data.
Query strategy optimization used hash function for mapping of data in database.
For parallel optimization task parallel library (TPL) is used to utilize CPU resources.
As their experimental result the optimized algorithm is over 100 times more efficient than the basic algorithm [6].
Bart at el. proposed a method for automatic emotion detection in suicidal notes using SVM classification and bootstrap re-sampling [7]. They considered 15 kinds of emotions for fine grained detection in suicide notes. They are forgiveness, abuse, pride, fear, happiness, hopefulness, sorrow, anger, thankfulness, blame, guilt, love, information, hopelessness, and instructions. Next figure shows the concept of the model which was proposed:



The dataset used was 1319 suicidal notes. The preprocessing of data is done by MBSP (Memory based shallow parser) which provide Lemmas and POS tags. The seven features are defined – Lemmas, Lemmas + POS tags, Pruned Lemmas + POS tags, Trigrams (three consecutive lemmas), WordNet Synsets (lexical database), SentiWordNet information (opinion mining resource), Subjectivity clues. Next step is for correction of misspelled words. They used fine grained (sentence level) sentiment identification with SVM classification followed by Bootstrap Resampling. F1 score is used as a measure of classification result, which is weighted average of precision and recall. Here precision measures the amount of false positives and recall measures the amount of false negative. In bootstrap resampling, various decision boundaries had considered instead of single one which is proposed by SVM classifier. It was used to determine which threshold maximizes F score. Bootstrap resampling experimented using SVM classifiers with 17 feature sets on the 15 emotions. The final output was obtained by aggregating the outputs of the best-performing classifier for each emotion [7].Haji Binali at el. has given a comparative analysis of various emotion detection methods and had given a generalized model as shown below [8]:



Data source identification involves selection of data sources which is to be analyzed. Text preprocessing includes stop words detection, tokenization, part-of-speech (pos) tagging, parsing (syntactic analysis), stemming and lemmatization. An emotion model is the means by which emotions are detected in text. It is built by selecting a technique in conjunction with a number of processes. There are four major categories of techniques that have been identified for building text based emotion detection models – corpus based model, machine learning based techniques, knowledge-based techniques, and hybrid techniques. Post-processing consists of some processes that are used in pre-processing. Results are the outputs of the various systems that are being studied/ applied. Evaluation of result is used to measure the effectiveness and efficiency of the model [8].

III. Comparative Analysis

S. No.	Author Names	Paper Title	Summary of the Paper
1.	Bart Desmet Veronique Hoste	Emotion Detection in Suicide Notes	(i) Automatic emotion detection in suicidal notes using SVM classification and bootstrap re-sampling. (ii) Considered 15 kinds of emotion with 17 feature sets. (iii) Dataset contained 1319 suicidal notes. (iv) F1 score is used as a measure of classification.
2.	Yu Zhang Zhuoming Li Fuji Ren Shingo Kuroiwa	Semi-automatic Emotion Recognition from Textual Input Based on the Constructed Emotion Thesaurus	(i) Emotion detection for Chinese natural language processing. (ii) Classified emotions into 12 categories with 5 levels (tag assignment -2 to 2). (iii) Collected 4800 words from various dictionaries. (iv) Used ICTCLAS Chinese lexical analyzer. (v) 80% accuracy in experimental result.
3.	Khin Phyu Shein Thi Thi Soe Nyunt	Sentiment Classification based on Ontology and SVM Classifier	(i) Ontology based sentiment classification. (ii) Ontology based on Formal Concept Analysis (FCA). (iii) Used SVM classifier. (iv) Experimental result provides positive/ negative classification.

4.	Prashant Raina	Sentiment Analysis in News Articles Using Sentic Computing	(i) Review of news article using sentic computing. (ii) The model contains semantic parser, sentic net database, and sentiment analyzer. (iii) Polarity measure is used to classify the sentences as positive, negative or neutral with score -1 to +1. (iv) 71.2% accuracy in experimental result.
5.	Yiming Zhao Kai Niu Zhiqiang He Jiaru Lin Xinyu Wang	Text Sentiment Analysis Algorithm Optimization & Platform Development in Social Network	(i) Chinese text analysis for social network. (ii) Based on SAMC algorithm. (iii) Basic SAMC algorithm is optimized by data structure optimization, query strategy optimization, parallel optimization. (iv) 100 times more efficient than basic SAMC algorithm.
6.	V. K. Singh R. Piryani A. Uddin P. Waila	Sentiment Analysis of Movie Reviews: A new Feature-based Heuristic for Aspect-level Sentiment Classification	(i) Aspect level sentiment classification for movie review. (ii) SentiWordNet based scheme is used for analysis. (ii) Based on 11 distinct features, movie rating is done as positive/negative.

IV. CONCLUSION

The objectives of this survey was to study various emotion detection techniques used in text based documents using Natural Language Processing. SVM classification is a popular technique which is used by the researchers. There is lot of application areas where works has been done but still there is a scope to have a wide range of application which will use emotion identification. The work so far done is mostly on the contents available on social media or on internet. The further extension of the existing methods can be applied to hand written documents which could be a challenging job.

REFERENCES

[1] Yu Zhang, Zhuoming Li, Fuji Ren and Shingo Kuroiwa, "Semi-automatic Emotion Recognition from Textual Input Based on the Constructed Emotion Thesaurus", IEEE Proceedings of NLP-KE 2005, P 571 – 576.
[2] Chandhana Surabhi M, "Natural Language Processing Future", Proceedings of International Conference on Optical Imaging Sensor and Security, Coimbatore, Tamil Nadu, India, July 2-3, 2013, IEEE proceeding.
[3] Prashant Raina, "Sentiment Analysis in News Articles Using Sentic Computing", 2013 IEEE 13th International Conference on Data Mining Workshops, P 959 – 962.
[4] V. K. Singh, R. Piryani, A. Uddin, P. Waila, "Sentiment Analysis of Movie Reviews: *A new Feature-based Heuristic for Aspect-level Sentiment Classification*", IEEE proceeding 2013, P 712 – 717.
[5] Khin Phyu Phyu Shein, Thi Thi Soe Nyunt, "Sentiment Classification based on Ontology and SVM Classifier", 2010 Second International Conference on Communication Software and Networks, IEEE proceeding, P 169 – 172.
[6] Yiming Zhao, Kai Niu, Zhiqiang He, Jiaru Lin, and Xinyu Wang, "Text Sentiment Analysis Algorithm Optimization & Platform Development in Social Network", 2013 Sixth International Symposium on Computational Intelligence and Design, IEEE proceeding, P 410 – 413.
[7] Bart Desmet, Veronique Hoste, "Emotion Detection in Suicide Notes", Expert Systems with Applications, Elsevier journal 2013, Volume 40 Issue 16, P 6351 – 6358.
[8] Haji Binali, Vidyasagar Potdar, "Emotion Detection State of the Art", ACM 2012, P 501 – 507.
[9] Harpreet Kaur, Veenu Mangat, Nidhi, "A Survey of Sentiment Analysis Techniques", International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) 2017, IEEE proceeding, P 921 – 925.
[10] Yasmina Douiji, Hajar Mousanif, I-CARE: Intelligent Context Aware System for Recognizing Emotions from Text", 2015 10th International Conference on Intelligent Systems: Theories and Applications: SITA.